Lineage-specific expansions of TET/JBP genes and a new class of DNA transposons shape fungal genomic and epigenetic landscapes

Lakshminarayan M. Iyer[a,1], Dapeng Zhang[a,1], Robson F. de Souza[b], Patricia J. Pukkila[c], Anjana Rao[d,2], and L. Aravind[a,2]

[a]National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894; [b]Microbiology Department, Biomedical Sciences Institute, University of Sao Paulo, 05508-900, Sao Paulo, Brazil; [c]Department of Biology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-3280; and [d]Sanford Consortium for Regenerative Medicine and La Jolla Institute for Allergy and Immunology, La Jolla, CA 92037

This contribution is part of the special series of Inaugural Articles by members of the National Academy of Sciences elected in 2008.

Contributed by Anjana Rao, December 17, 2013 (sent for review September 20, 2013)

TET/JBP dioxygenases oxidize methylpyrimidines in nucleic acids and are implicated in generation of epigenetic marks and potential intermediates for DNA demethylation. We show that TET/JBP genes are lineage-specifically expanded in all major clades of basidiomycete fungi, with the majority of copies predicted to encode catalytically active proteins. This pattern differs starkly from the situation in most other organisms that possess just a single or a few copies of the TET/JBP family. In most basidiomycetes, TET/JBP genes are frequently linked to a unique class of transposons, KDZ (Kyakuja, Dileera, and Zisupton) and appear to have dispersed across chromosomes along with them. Several of these elements typically encode additional proteins, including a divergent version of the HMG domain. Analysis of their transposases shows that they contain a previously uncharacterized version of the RNase H fold with multiple distinctive Zn-chelating motifs and a unique insert, which are predicted to play roles in structural stabilization and target sequence recognition, respectively. We reconstruct the complex evolutionary history of TET/JBPs and associated transposons as involving multiple rounds of expansion with concomitant lineage sorting and loss, along with several capture events of TET/JBP genes by different transposon clades. On a few occasions, these TET/JBP genes were also laterally transferred to certain Ascomycota, Glomeromycota, Viridiplantae, and Amoebozoa. One such is an inactive version, calnexin-independence factor 1 (Cif1), from *Schizosaccharomyces pombe*, which has been implicated in inducing an epigenetically transmitted prion state. We argue that this unique transposon-TET/JBP association is likely to play important roles in speciation during evolution and epigenetic regulation.

methylcytosine | fungal evolution | DNA modification | genomic association

Intragenomic conflicts with diverse mobile elements have played a critical role in shaping cellular genomes (1). In eukaryotes, genomic rearrangements mediated by transposons in germ-line cells promote reproductive isolation of organisms and, accordingly, are implicated in speciation events (2, 3). The mobility of transposons has the potential to disrupt genes by insertion as well as to create new genes or to resurrect inactive ones. Indeed, genomic analyses have revealed that transposons are an important source for both regulatory elements and new DNA-binding domains in transcription factors and chromatin proteins across the tree of life (1, 4). Organisms show a diverse array of strategies to counter mobile elements, supporting the idea of a constant arms race between them and genomes. Some of these strategies, such as transcriptional silencing by chromatin modifications (e.g., histone methylation, DNA methylation) and posttranscriptional silencing using components of the RNAi machinery, are widespread across eukaryotes (5). In contrast, processes such as DNA elimination in the macronuclei of ciliates show a restricted phyletic distribution (6). Transposons have, in turn, evolved a variety of adaptations that help them survive in host genomes (7). For example, eukaryotic transposons regulate their interaction with chromatin with adaptations, such as histone-recognition domains (e.g., transposase-associated chromodomains and PHD fingers) to interact with or introduce (e.g., SET domain methyltransferase in certain Mariner elements) epigenetic modifications (8). Furthermore, the initially antagonistic interactions between transposons and the host genome can be eliminated over time via "domestication" of the transposon, wherein the transposon is incorporated as one or more cellular genes, with concomitant attenuation or loss of transposition. Such domesticated transposons might play important roles as catalysts of controlled genome rearrangements: The animal recombination activating gene 1 (RAG1) recombinase in immune receptor diversification is derived from Transib elements, the yeast mating-type switch nuclease is derived from mobile homing endonucleases, and the ciliate PiggyBac elements are involved in macronuclear DNA elimination (9–11).

We previously described a class of DNA transposons in basidiomycete fungi, namely, *Coprinopsis* and *Laccaria*, that encoded a predicted transposase and a protein of the TET/JBP family (12). The TET/JBP proteins are 2-oxoglutarate-Fe^{2+}-dependent dioxygenases (2OGFeDO) that catalyze hydroxylation of the exocyclic carbon at the 5 position of pyrimidines in DNA (13–15).

Significance

5-Methylcytosine in DNA of eukaryotes, such as humans, is an important epigenetic mark. The recently characterized TET/JBP enzymes generate oxidized derivatives of methylcytosine, such as hydroxy-, formyl-, and carboxymethylcytosine in mammals, which serve as further epigenetic marks or intermediates for demethylation. Unlike animals, which contain one to three TET genes, fungi, such as mushrooms and rusts, display lineage-specific expansions with numerous TET/JBP genes, which are often associated with a unique class of transposable elements. We present evidence that expansion and turnover of these elements and associated TET/JBP genes play important roles in genomic organization, epigenetics, and speciation of fungal lineages, especially basidiomycetes (mushrooms, rusts, and smuts). Domesticated versions of these transposons might also participate in genome rearrangements or repair in humans.

Author contributions: L.M.I., D.Z., A.R., and L.A. designed research; L.M.I., D.Z., and L.A. performed research; L.M.I., D.Z., R.F.d.S., P.J.P., and A.R. contributed new reagents/analytic tools; L.M.I., D.Z., R.F.d.S., and L.A. analyzed data; and A.R. and L.A. wrote the paper.

The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

[1]L.M.I. and D.Z. contributed equally to this work.

[2]To whom correspondence may be addressed. E-mail: arao@liai.org or aravind@ncbi.nlm.nih.gov.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1321818111/-/DCSupplemental.

Animal TET proteins successively oxidize methylcytosine (mC), generating hydroxymethylcytosine, formylcytosine, and carboxycytosine, collectively referred to here as oximC. In contrast, the kinetoplastid JBP proteins catalyze hydroxylation of thymine, which is further glycosylated to give the hypermodified base, β-D-glucopyranosyloxymethyluracil or base J (14, 16). Sequence and phylogenetic analyses suggested that the TET/JBP family originated in bacteriophages. They were acquired from phages or bacteria and recruited as generators of epigenetic marks by eukaryotes on at least three independent occasions (15). Two distinct phyletic patterns of TET/JBP proteins are observed across all organisms. In animals, *Acanthamoeba*, certain algae (e.g., *Coccomyxa*), *Naegleria*, kinetoplastids, and certain bacteria and phages, there is a single or just a few copies of TET/JBPs, with the main evolutionary trends being lateral transfer from bacterial/bacteriophage sources and a tendency for vertical inheritance. In contrast, the second pattern, first observed in *Coprinopsis* and *Laccaria*, is one of massive expansions, often with 10 or more copies frequently coupled with transposons. Strikingly, unlike typical transposon-associated genes, where only a few of the many copies are active in a genome (17), analysis of the TET/JBP domains in the above mushrooms showed that the majority of them are predicted to be catalytically active (12).

In eukaryotes other than kinetoplastids, there is a strong phyletic congruence with the presence of DNMT1 orthologs (*SI Appendix*, Table S1), suggesting that in most of these cases, the TET/JBP proteins are likely to oxidize mC, as exemplified by several studies on the metazoan TET proteins (14, 15). Consistent with this observation, it was shown that *Coprinopsis* contains oximC that is enriched both at transposons coding for TET/JBP proteins as well as at other transposons and repetitive DNA. Hence, TET/JBP proteins are likely to catalyze these modifications from mC generated by the two DNMT1 orthologs. Since the initial description of the fungal TET/JBP family, a large number of new genomes have been sequenced across the fungal tree. Preliminary examination of these genomes revealed that comparable expansions of TET/JBP and associated transposons are rife in several fungal species (15). Hence, we sought a better understanding of these expansions by analysis of their genomic associations, domain architectures of the encoded proteins, and their phyletic spread.

Results and Discussion

TET/JBP Genes Show Extensive Lineage-Specific Expansion in Fungi. To understand the complete extent of the expansion of TET/JBPs in fungi, we comprehensively collected them using sensitive profile search algorithms (PSI-BLAST and JACKHMMER), followed by confirmation with profile-profile comparisons with the HHpred program. The resulting phyletic patterns (*SI Appendix*, Table S1) point to the presence and expansions of TET/JBPs in representatives from all three major clades of Basidiomycota, namely, Pucciniomycotina (rusts), Ustilagomycotina (smuts), and Agaricomycotina (mushrooms), and infrequently in Ascomycota (Pezizomycotina and Taphrinomycotina) and Glomeromycota (arbuscular mycorrhizal plant endosymbionts). Although most Agaricomycetes showed large expansions of TET/JBPs (>10 paralogs) some representatives, such as *Rhizoctonia*, *Postia*, and *Fomitiporia*, contained very few (less than three paralogs). In certain cases, we found that even sister species differed greatly in their TET/JBP counts (e.g., *Sporisorium reilianum* and *Ustilago hordei* show expansions of TET/JBP, whereas *Ustilago maydis* lacks them entirely). Among pucciniomycetes, large expansions are seen in plant pathogens, such as *Puccinia* and *Melampsora larici-populina* (100–200 paralogs), but they are seen in far fewer copies (four to five paralogs) in the related yeasts, such as *Rhodotorula*. Currently, in Ascomycota, TET/JBPs are only found sporadically in few members of Pezizomycotina, with expansions in the grape dead-arm fungus *Eutypa lata* and the bat pathogen *Geomyces destructans* (*SI Appendix*, Table S1). Interestingly, we observed that the fission yeast *Schizosaccharomyces pombe* protein Cif1 (SPCC364.01), which induces an epigenetically transmitted, prion state (18), is a catalytically inactive version of the fungal TET/JBPs.

The fungal TET/JBPs were found to form a monophyletic clade along with representatives in chlorophytes, such as *Volvox* and *Chlamydomonas*, the land plant *Physcomitrella* (a bryophyte moss), and the amoebozoan *Acanthamoeba* (15), which were distinct from all other eukaryotic TET/JBP proteins (15). Sequence and phylogenetic analysis revealed four prominent clades of fungal TET/JBPs (Fig. 1). The Pucciniomycete-like clade is dominated by sequences from *Puccinia* and *Melampsora*, and additionally contains a cluster of proteins from Agaricomycetes. These are usually characterized by a conserved HxT (where x is any residue) signature in the context of the C-terminal Fe-chelating histidine of the TET/JBP double-stranded β-helix (DSBH) (12). The chlorophyte and land plant versions are likely to have been transferred from this clade (2). The *Coprinopsis*-like clade, containing sequences mainly from *Coprinopsis* and other Agaricomycetes, is characterized by a conserved Rx[3–5]HxD signature in the context of the N-terminal Fe-chelating positions of the DSBH (3). The *Auricularia*-like clade, dominated by sequences from *Auricularia delicata*, also contains sequences from several other Agaricomycotina. This clade is characterized by a conserved HxN signature in the context of the C-terminal Fe-chelating histidine of the DSBH (4). The Ascomycota-like clade contains sequences from ascomycetes, such as *E. lata* and *Schizosaccharomyces*, as well as sequences from Ustilagomycotina (e.g., *S. reilianum* and *U. hordei*) and Glomeromycota (*Rhizophagus*). Several of these sequences have a characteristic HxG signature associated with one of the strands N-terminal to the core DSBH fold.

A striking aspect of the tree is that the TET/JBP members from each fungal species tend to fall into one or a few strongly supported clusters (bootstrap support >85%) entirely composed of monospecific representatives (Fig. 1). For example, 33 (94%) of 35 *Coprinopsis* sequences in the tree are present in the *Coprinopsis* clade and fall into two monospecific clusters. Beyond the monospecific clusters, there are hints of a vertical phylogenetic signal: Monospecific clusters of sister species, such as *Coprinopsis* and *Laccaria* or *Melampsora* and *Puccinia*, tend to form strongly supported higher order clades (Fig. 1 and *SI Appendix*, Fig. S1). The presence of TET/JBP proteins in all major groups of Basidiomycota suggests that they were first acquired in the common ancestor of this lineage. Moreover, the presence of sequences from the same species in multiple major TET/JBP clades suggests that the ancestral basidiomycete was likely to have encoded several copies of TET/JBP proteins. Together, these observations suggest that right from inception, repeated cycles of lineage-specific expansion of the TET/JBP proteins, accompanied by episodes of sorting with inheritance of certain versions and loss of others, routinely occurred during speciation in Basidiomycota (*SI Appendix*, Fig. S1). In some lineages, upon sorting, there appears to have been no expansions leading to partial or complete loss of the TET/JBP complement. The limited and sporadic presence in Ascomycota, Glomeromycota, and Mucoromycotina is suggestive of lateral transfer of TET/JBPs from basidiomycetes, followed by occasional expansions in a subset of these organisms (*SI Appendix*, Fig. S2). All versions from Viridiplantae and *Acanthamoeba* were nested within the fungal versions, suggesting that they were also laterally transferred from fungi to those organisms (*SI Appendix*, Fig. S1).

KDZ Class of Transposons. To understand the transposons that are associated with TET/JBP genes in *Coprinopsis* and *Laccaria* (12), and their possible connection to the lineage-specific expansions of those genes, better, we systematically investigated their phyletic spread, genomic organization, and domain architectures of the encoded proteins. Profile-based sequence searches with the

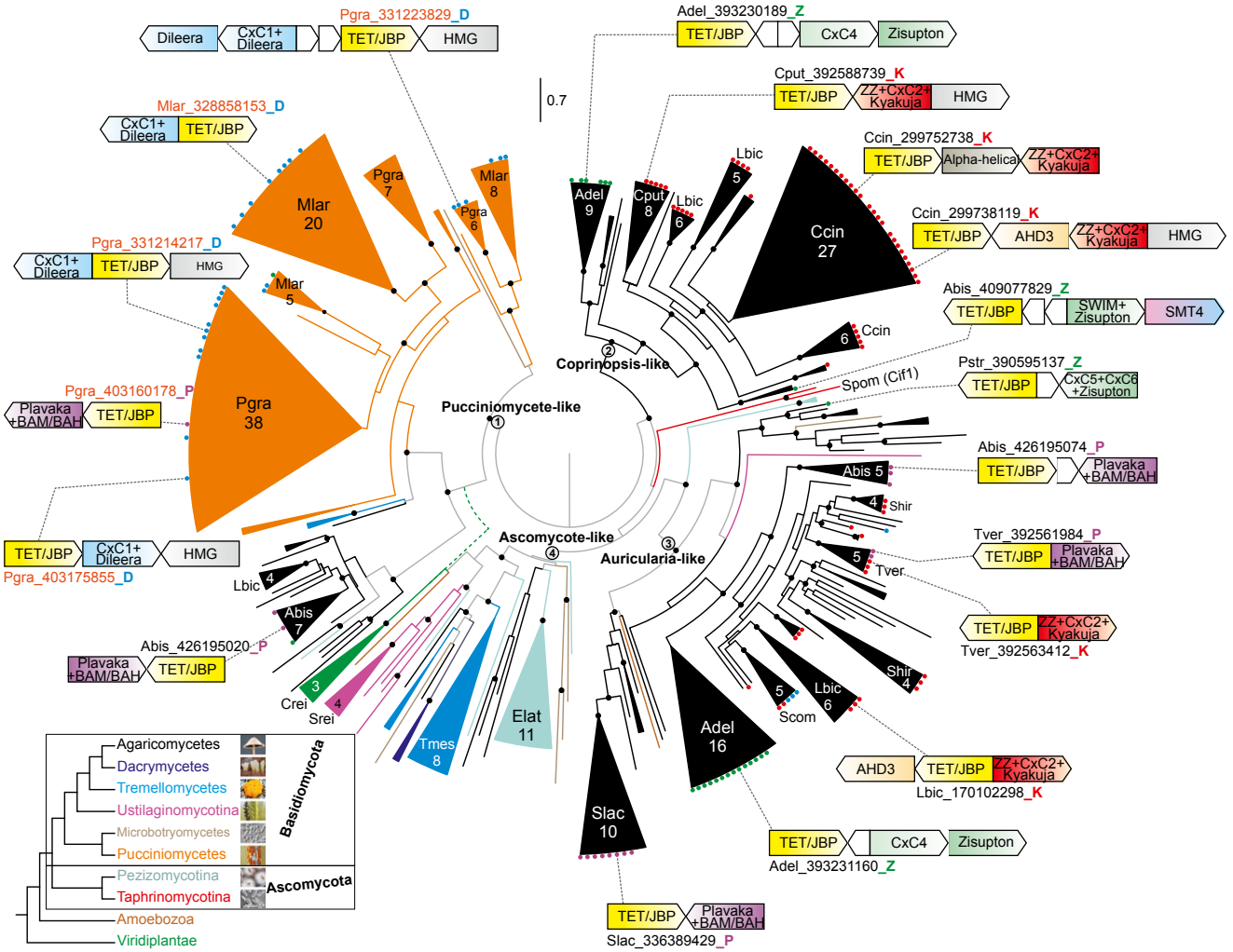

Fig. 1. Phylogenetic tree shows lineage-specific expansions of TET/JBP proteins in different fungi. Branches and names are colored differentially based on their lineages. Clades entirely composed of monospecific representatives are collapsed into triangles and labeled with the species abbreviation and number of sequences in the clade. Nodes supported by bootstrap values >85% are marked with black circles. Each colored dot shown at the edge of the collapsed clades indicates a single association between a TET/JBP protein and a transposase, where Kyakuja (K) is shown in red, Dileera (D) in blue, Zisupton (Z) in green, and Plavaka (P) in purple. Genomic structures of representative examples of these associations are shown around the tree and are labeled with the species abbreviation, Genbank index number (gi) of the TET/JBP gene, and the type of associated transposase (i.e., K, D, Z, or P) in the association. Genes are shown as arrows pointing from the 5′ end to the 3′ end, with the name and domain architecture within. A fully expanded tree with individual branch labels is provided in *SI Appendix*, Fig. S1. Species abbreviations are provided in *SI Appendix*, Table S1.

distinct transposase retrieved 1–211 copies in genomes of fungi, animals, chlorophytes, and stramenopiles (*SI Appendix*, Table S1). A phylogenetic tree of the transposase domains showed that these elements form three distinct divergent clades (*SI Appendix*, Fig. S3). The first of these included those that we had initially found to be associated with TET/JBP genes in *Coprinopsis* and *Laccaria*; we named these the Kyakuja (mushroom in Sanskrit) elements. The remaining two were identified as related transposons for which we initially did not find any associations with TET/JBP genes (12). We named the second clade Dileera (another word for mushroom in Sanskrit). The third clade was recently rediscovered (19) and shown to be active transposons in fishes called the Zisuptons (the Zisupton clade). Accordingly, we collectively refer to these transposons as the KDZ class. Their phyletic patterns and relationships strongly suggest mobility between major eukaryotic lineages, followed by repeated expansions in certain lineages in a pattern similar to that of the TET/JBP genes in fungi (Fig. 1 and *SI Appendix*, Fig. S3).

Although we had proposed that the catalytic domain of the KDZ transposases contained an RNase H fold, their active site and catalytic mechanism remained unclear (12). Our current analysis using profile-profile and secondary structure comparisons of KDZ transposons helped to define the core RNase H fold catalytic domain of these transposases correctly and to identify the catalytic residues that are prototypic features of RNase H-fold transposases, namely, the acidic residues after strands 1 and 4 and a glutamate in helix 3 (Fig. 2 and *SI Appendix*, Figs. S4 and S5). Equivalent conserved catalytic residues coordinate a divalent cation in Mariner-like, Hermes-like, Transib-like, and mutator-like element (MULE) transposases and retroviral integrases (20). KDZ transposase domains are uniquely characterized by a dyad of metal-chelating residues (the CxH signature) occurring after strand 2 in the RNase H-fold domain, which are predicted to form a Zn cluster along with a dyad of cysteines just N-terminal to the first strand of the RNase H fold (Fig. 2*A* and *SI Appendix*, Figs. S4 and S5). This structural Zn cluster is likely to be positioned on the face

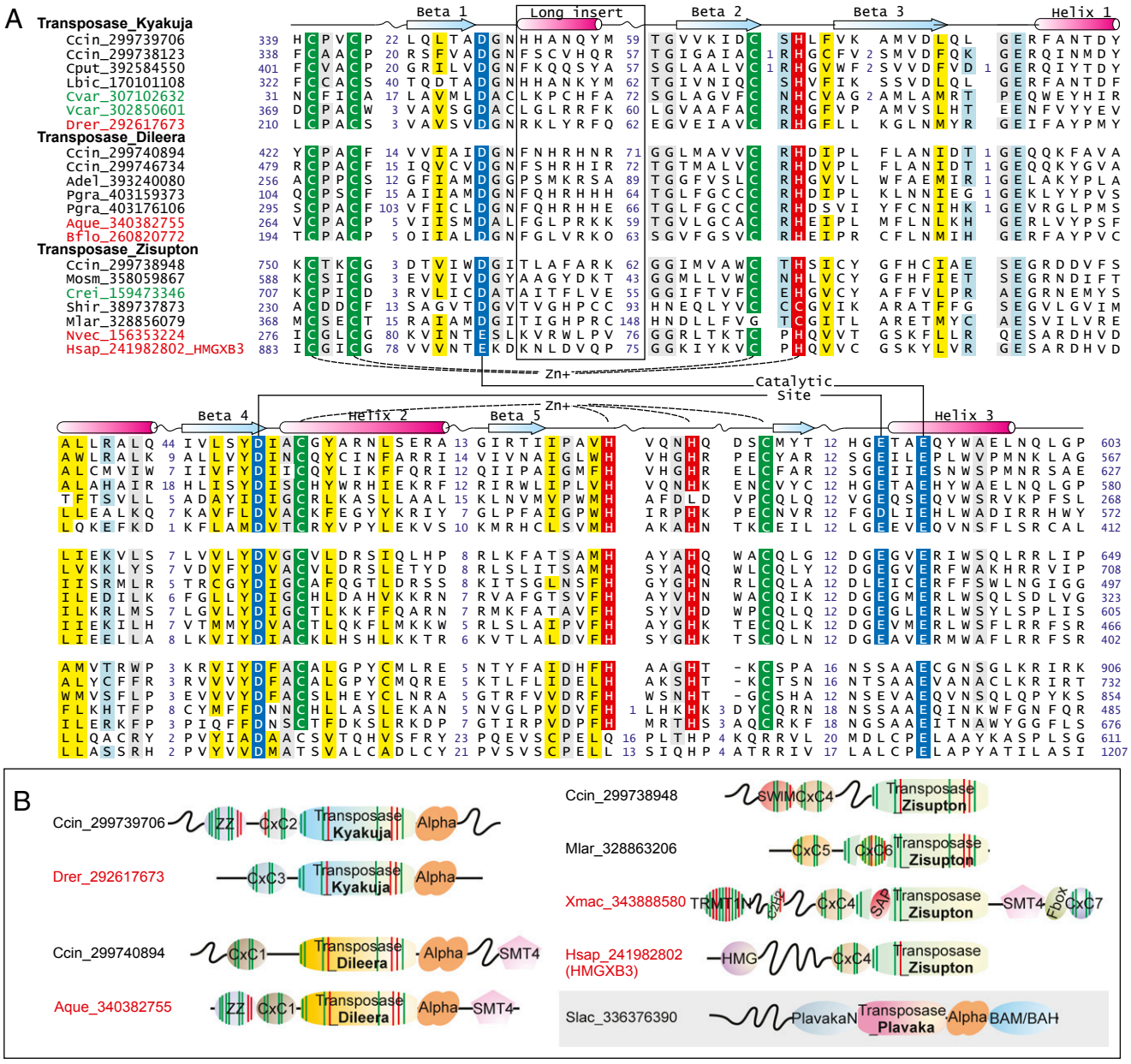

Fig. 2. (*A*) Multiple sequence alignment of KDZ transposases. Proteins are labeled with their species abbreviations and GenBank index numbers. Species abbreviations are provided in *SI Appendix*, Table S1. Conserved Zn-binding residues and catalytic residues of transposases are indicated. The coloring is based on 85% consensus using the following scheme: polar residues (CDEHKNQRST) shaded light blue, hydrophobic (ACFILMVWY) residues shaded yellow, and small (ACDGNPSTV) residues shaded gray. Catalytic residues are colored dark blue. Conserved cysteine and histidine residues predicted to be involved in coordinating a Zn ion are colored green and red, respectively. (*B*) Domain architectures of representative KDZ and Plavaka transposases. In each domain, the conserved Zn-chelating residues, cysteine and histidine, are shown as vertical green and red lines.

opposite to the active site of the transposase. The KDZ transposase RNase H-fold domain is also characterized by a giant insert between strands 1 and 2 of the core fold, which is stabilized, in part, by metal-chelating residues in Kyakuja and Dileera elements occurring both within and beyond this insert (Fig. 2 and *SI Appendix*, Fig. S5). In Zisupton elements from animals, this insert contains the DNA-binding SAP domain (19), whereas those from fungi contain a dyad of zinc ribbons (CXC6; Fig. 2*B* and *SI Appendix*, Fig. S6). A well-studied transposon family with a comparable insert between strands 1 and 2 is Mariner (Mos1), where the crystal structure shows that the insert is involved in making specific contacts with the target DNA (21). This feature of the Mariner RNase H-fold domain, together with presence of the SAP domain in animal Zisuptons within this insert, points to a comparable DNA-binding role for this region across the KDZ superfamily of transposases. The variability, both in terms of inserted domains and sequence conservation within and between the three major clades, suggests that this insert might play a role in recognition of target sites among these transposons.

Identification of the RNase H-fold catalytic residues in these transposase domains allows us to distinguish active from inactive transposases objectively. On the basis of these residues, a notable fraction of these KDZ transposases in fungi were predicted to be catalytically active (Fig. 2*A* and *SI Appendix*, Fig. S4). Interestingly, the apparently domesticated transposase domain of the vertebrate HMGBX3, which is related to the Zisupton-type

transposase (19), is also predicted to be a catalytically active endoDNase (Fig. 2*A* and *SI Appendix*, Fig. S4), suggesting that it could be involved in genome repair or reorganization comparable to that of the RAG1 transposase domain. This high fraction of active transposase domains is reminiscent of the active DNA transposons in the micronucleus of the ciliate *Oxytricha trifallax* (11). It is notably different from retrotransposons and retrovirus-like and DIRS-1–like retrotransposons (22), as well as DNA elements, such as Tigger, where the majority of copies tend to have inactive domains (17). This suggests that the transposases of at least a subset of KDZ elements might have functional consequences for the genomes harboring them, comparable to the *O. trifallax* elements involved in genome rearrangements during macronuclear maturation (11).

Beyond the transposase domain, most of the KDZ transposases possess a Zn-chelating domain with four conserved cysteine/histidine residues N-terminal to the catalytic domain (labeled as CxC1–5 with a number in Fig. 2*B* and *SI Appendix*, Figs. S7–S11). A subset of these elements contains further N- or C-terminal Zn-chelating domains, such as the TRMT1N-like domain, SWIM, C2H2 fingers, and CXC7 in Zisuptons and the ZZ finger in Kyakujas (Fig. 2*B* and *SI Appendix*, Figs. S12–S14). These domains could help recognize specific DNA sequences or chromatin proteins during integration. Several Zisupton transposases contain C-terminal peptidase domains of the deubiquitinating (DUB) Ulp1/Smt4 or OTU-like superfamilies of the papain-like fold (19) (Fig. 2*B*). Although no peptidase domains were ever found fused to the transposase catalytic domain in the Kyakuja elements, a subset of the Dileera elements from fungi was found to contain Ulp1/Smt4-like peptidases at their C termini (Fig. 2*B*). Given that cellular counterparts of these peptidase domains function as DUBs (23), it is conceivable that the transposon-encoded versions help cleave sumoylated or ubiquitinated chromatin or self-encoded proteins as part of the regulation of their integration. Alternatively, as with RNA viral or retroviral peptidases that process viral polyproteins, they might cleave the large transposase polypeptide (24).

Zisupton elements are believed to be flanked by long (>100 nt) terminal inverted repeats (TIRs) (19). However, our analysis of the fungal Zisuptons, as well as the Dileera and Kyakuja elements, did not reveal any widely conserved TIRs greater than 10 nt. If longer TIRS indeed existed, they might have been lost due to divergence, with the transposase domain being domesticated independent of repeats. Alternatively, most of these elements might use mechanisms distinct from recognition of TIRs for their DNA rearrangements (recognition of chromatin proteins or modified DNA containing oximC). A third possibility is that such repeats exist only in a single or small subset of the elements in any given genome (i.e., "fully functional" copies).

Association Between KDZ Superfamily Transposons and TET/JBP Genes. In fungi that possess both TET/JBP genes and KDZ transposons, we observed a general positive correlation between their numbers (*SI Appendix*, Table S1). Hence, we investigated if there was a genomic association between the two across these fungi, just as we had observed earlier in *Coprinopsis* and *Laccaria.* Accordingly, we systematically mapped chromosomal locations of TET/JBP and KDZ transposase genes and identified instances where they co-occur as neighboring genes. Further examination of these neighborhood associations allowed us to identify elements of similar length combining TET/JBP and KDZ transposase genes, which appear to have proliferated and dispersed to multiple chromosomal sites in a given organism. As in the case of *Coprinopsis*, where the complete Kyakuja elements additionally code for an HMG domain protein and usually one other ORF, even in these other fungi, the elements coded for one or two additional ORFs (Fig. 1 and *SI Appendix*, Figs. S15 and S16 and Table S1). Simulations of co-occurrence of genes in neighborhoods of similar overall length at different chromosomal locations showed that the probability of these occurring by chance alone was extremely low ($P <= 10^{-4}$; details are provided in *SI Appendix*, Fig. S17). Further, in the overall phylogenetic tree of TET/JBP proteins, we observed that those associated with neighboring transposase genes tended to group together, hinting that the lineage-specific expansion and association with a transposase are related (Fig. 1). Mapping of the transposase associations onto the phylogenetic tree of the TET/JBP proteins also indicated that all three major clades, Kyakuja, Dileera, and Zisupton, have associated with TET/JBP genes on different occasions. For example, in *Coprinopsis*, the association is entirely with Kyakuja. In *Auricularia*, the dominant association is with Zisupton, and in *Melampsora* and *Puccinia*, the main association is with Dileera.

We then tested the congruence between the phylogenetic trees of TET/JBPs and their respective associated transposase proteins. We used two methods for comparing phylogenetic trees that either produce a topological alignment between the pairs of trees (25) or perform a nodal comparison between trees along with a randomization analysis to test whether the similarity between two trees is due to chance (26). These analyses showed that pairs of TET/JBP and associated transposase trees had high topological congruence (70–92%), and their nodal congruence was 2–6 SDs from that of an average tree expected by chance alone (Fig. 3 and *SI Appendix*, Fig. S18). These observations strongly support the comobility of TET/JBP genes with different transposase genes from the KDZ superfamily representatives in the course of their proliferation. In addition to the three KDZ-type transposons, we found a fourth distinctive type of transposase, which is associated with the TET/JBP genes, especially in *Serpula lacrymans* and *Fomitopsis pinicola*. Sequence analysis revealed that it contains a unique version of the RNase H catalytic domain but shares none of the distinct structural features of the KDZ superfamily (Fig. 2*B* and *SI Appendix*, Fig. S19). These transposases are also distinguished by a unique version of the histone-binding BAM/BAH domain at the C terminus (*SI Appendix*, Fig. S20). We named this family of transposons the Plavaka elements (jumper in Sanskrit). Unlike the KDZ class, support for their association with TET/JBP genes is currently limited; nevertheless, this observation does raise the possibility that there might be other rare transposon-TET/JBP associations in fungal genomes.

Based on the above phyletic patterns (*SI Appendix*, Table S1) and trees (Figs. 1 and 3 and *SI Appendix*, Fig. S3), we propose that the association between TET/JBP and KDZ transposons probably emerged in the ancestor of Basidiomycota itself. This was followed by multiple capture events, among which the major ones can be parsimoniously reconstructed as follows: (*i*) capture by Dileera transposons in Pucciniomycetes; (*ii*) capture by Kyakuja transposons at the base of Agaricomycotina of TET/JBP genes belonging to both *Coprinopsis*-like and *Auricularia*-like clades; (*iii*) major secondary capture of TET/JBPs of both the *Coprinopsis*-like and *Auricularia*-like clades by Zisuptons in *A. delicata* and of *Auricularia*-like TET/JBPs by Dileeras in *Schizophyllum commune*; and (*iv*) several sporadic secondary captures of TET/JBPs, mainly of the *Auricularia* clade, by Plavaka in *Serpula*, *Fomitopsis*, *Agaricus*, and *Tremetes* species (Fig. 1 and *SI Appendix*, Figs. S1 and S2). Such sporadic events also include the recent capture of a Pucciniomycete-like TET/JBP [National Center for Biotechnology Information (NCBI) gi: 403160178] by a Plavaka transposase, whereas most of its paralogs associate with Dileera. The link between transposases and TET/JBP genes appears to have been lost in Ustilagomycotina and Tremellomycetes, where these transposons are entirely absent. Likewise, a version of the Pucciniomycete-like TET/JBP transferred to the Agaricomycetes shows no association with transposons in the latter (Fig. 1 and *SI Appendix*, Fig. S1). Intriguingly, despite expansions in certain Ascomycota and Glomeromycota, there

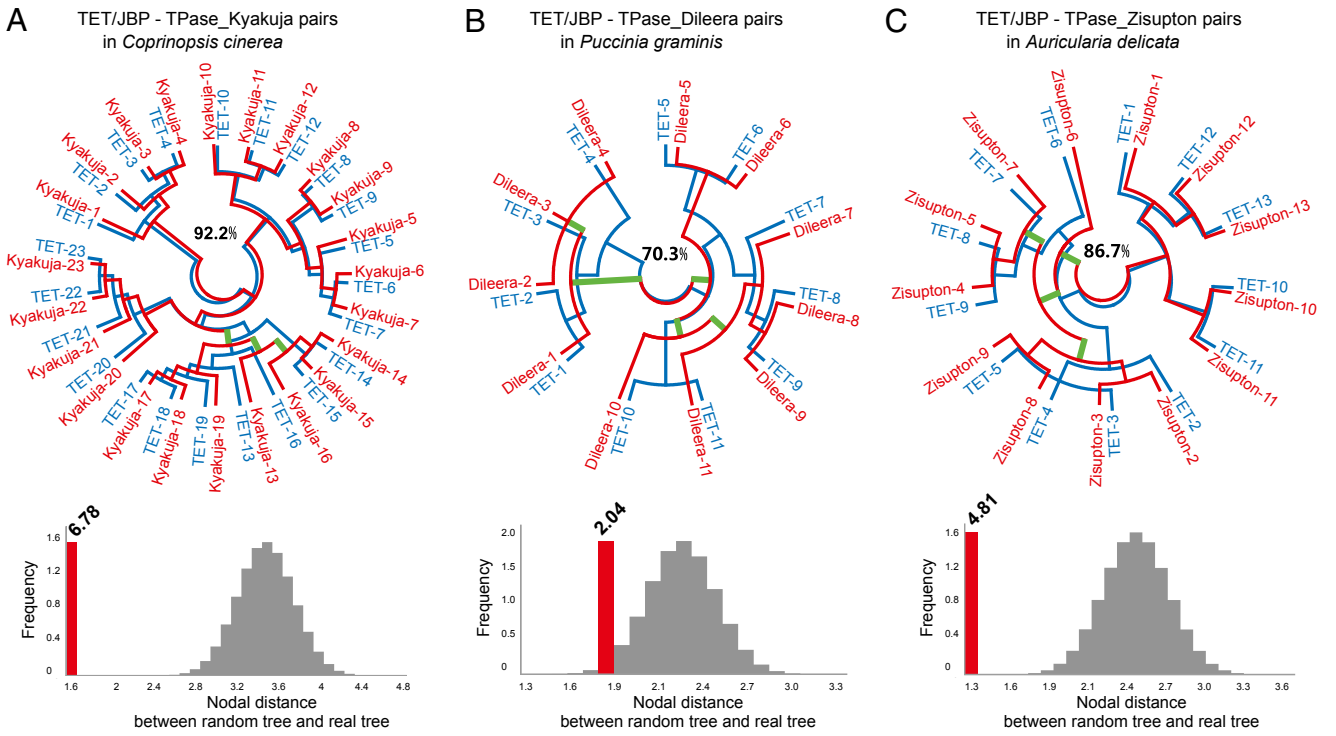

Fig. 3. Evolutionary associations of fungal TET/JBP and KDZ transposases. On the top are shown the topological alignments of the phylogenetic trees of TET/JBP and KDZ transposases with the TET/JBP tree colored in blue and transposase tree in red and branches showing topological mismatch in green. TET/JBP and KDZ transposase genes predicted to be derived from the same mobile element are assigned the same number shown next to the leaf label (a complete description of the leaves is provided in *SI Appendix*, Fig. S18). The red bar in the bottom graph is the actual recovered nodal distance positioned with respect to the normally distributed nodal distances of 1000 randomly generated trees from the same set of terminal leaves and is labeled on top using the number of SDs from the mean distance of the random trees.

are no associations between TET/JBP genes and KDZ transposons in these organisms. This suggests either that transfers of TET/JBP genes to these fungi happened independent of the associated transposase gene or that the latter was rapidly lost upon transfer.

Functional Implications and General Conclusions. Independent expansion of TET/JBP genes and associated DNA transposons in several fungi is in stark contrast to those in animals, which show one to three vertically inherited TET/JBP genes. Although in both animals and fungi, the presence of active TET/JBPs is correlated with DNMT1 (two copies of DNMT1 are present in most basidiomycetes; *SI Appendix*, Table S1), the rest of their methylation systems show notable differences. In fungi, DNA methylation is widely used to repress or inactivate duplicate copies of genes, thus serving as a mechanism of dosage compensation (27, 28). Unlike animals, fungi lack prominent gene body methylation and tend to concentrate their methylation to repetitive sequences (29). Concomitantly, they have lost TAM/MBD proteins, which bind completely methylated CpG sites, and the canonical SAD/SRA proteins, which recognize hemimethylated sites (30). This suggests that mechanisms related to removal of mC marks are also likely to be closely associated with parts of the genome enriched in repeats and mobile elements. In this context, it might be advantageous for certain transposable elements to bear TET/JBP genes to reverse or modulate the host organism's methylation-based silencing mechanisms (*SI Appendix*, Fig. S2). Analysis of Shannon entropy plots derived from TET/JBP alignments demonstrated that residues associated with the active site and key structural positions of the core DSBH show the lowest entropy throughout fungi, whereas all other positions are less constrained (*SI Appendix*, Fig. S21). This indicates that several of the transposon-associated TET/JBP genes are active across fungi, suggesting that the host genome might have developed a mutualism with transposable elements bearing TET/JBP genes to facilitate generation or resetting of epigenetic marks through an oxidative pathway.

One interesting possibility emerging from this study is the potential role for TET/JBP genes and associated transposons in fungal speciation. For example, 39% of the *Coprinopsis cinerea* genome is syntenic with *Laccaria bicolor* (31), but there is hardly any synteny with respect to their TET/JBP and KDZ transposase genes. Similarly, the closely related Pucciniomycetes *Melampsora* and *Puccinia* show extensive breakdown of genomic synteny (32), which again appears to be correlated with the massive independent expansions of TET/JBPs and associated Dileera elements. Indeed, in several fungi, including *Coprinopsis,* extensive karyotype variation from chromosomal rearrangements either in the course of development or between strains has been observed previously (33, 34). It is conceivable that the active transposases of KDZ elements have a role in the process of karyotype variation, leading to loss of synteny and consequent speciation. Importantly, vertebrate genes derived from KDZ transposons, such as the HMG domain protein HMGBX3, could play a similar role in genomic reorganization (e.g., in somatic cells) or DNA repair events. A precedent for this is offered by the mammalian domesticated SET domain-containing Mariner element that suppresses chromosome translocations (35). Moreover, evidence from well-assembled genomes (e.g., *Coprinopsis*) indicates that KDZ elements and TET/JBP elements tend to cluster in subtelomeric regions, where duplicated paralogous genes and retroposon-related sequences are overrepresented, in contrast to a less biased distribution of other transposons along the chromosome (31). Thus, generation of oximC epigenetic marks, along with rearrangements triggered by transposases, could alter

chromatin organization in these regions and affect expression of neighboring genes. Related TET/JBP genes from those fungi and Viridiplantae, in which they are uncoupled from KDZ or Plavaka transposases, might be seen as completely "domesticated" forms incorporated into the DNA methylation-dependent epigenetic machinery, comparable to animal TETs.

In *Schizosaccharomyces* and the moss *Physcomitrella*, inactive TET/JBPs are conserved over long evolutionary time, as indicated by the presence of distinct paralogs or conservation between species (e.g., *S. pombe*, S. *japonicum*); hence, they appear to have been selected for noncatalytic regulatory functions in these organisms. *S. pombe* TET/JBP, Cif1, an intriguing nucleolus-localized protein, induces an epigenetic state transmitted like a prion in the cytoplasm to provide resistance against protein misfolding stress (18). However, once established, Cif1 is not required for perpetuation of the prion state. Cif1 could possibly bind DNA to induce this state but plays no role thereafter in transmitting it. *S. pombe* lacks DNMT1, and it is unknown if its DNMT2 can methylate DNA in cells; hence, it is uncertain if DNA binding by Cif1 depends on methylation. Nevertheless, it is possible that it shares an epigenetic function with active TET/JBPs from other fungi, albeit through a noncatalytic DNA-binding mechanism. A subset of chlorophytes and basidiomycetes with TET/JBP genes also encodes the 5-hydroxylmethyluracil pyrophosphorylase, which we recently identified as being involved in the synthesis of hypermodified thymines in bacteriophages like SP10 and phi-W14 (15). These pyrophosphorylases could use pyrophosphorylated 5-hydroxymethyluracil derived from an intermediate generated by the TET/JBPs to synthesize hypermodified thymines. We also recovered methyltransferases derived from a unique group of prokaryotic restriction-modification systems typified by the pneumococcal DpnII system predicted to catalyze another DNA modification, N6 adenine methylation, in Glomeromycota, Mucoromycota, and Euchytrids (*SI Appendix*, Table S1). Thus, fungal lineages might possess a richer modified DNA landscape than is currently known (15).

In conclusion, these findings open the door for discovery and exploration of previously unexpected facets of fungal epigenetics dependent on oxidized mC species and chromosome dynamics dependent on recruitment of selfish elements.

Materials and Methods

Sequence Searches and Domain Analysis. Iterative sequence profile searches were performed using the PSI-BLAST and JACKHMMER programs against the NCBI nonredundant protein database (36, 37). Multiple sequence alignments were built using Kalign2 or Muscle, followed by manual adjustments based on profile-profile comparisons, secondary structure, and structural alignments (38). Similarity-based clustering for both classification and culling nearly identical sequences was performed using the BLASTCLUST program (ftp://ftp.ncbi.nih.gov/blast/documents/blastclust.html). The HHpred program was used for profile-profile comparisons (39). Secondary structures were predicted using the JPred program (40). For previously characterized domains, the Pfam database was used as a guide. Clustering with BLASTCLUST, followed by multiple sequence alignment and further sequence profile searches, was used to identify novel domains not present in the Pfam database (41). Structural visualization and manipulations were performed using the PyMol program (www.pymol.org).

Genomic Gene Neighborhood Analysis. To analyze gene neighborhoods, we extracted seven upstream and downstream neighbors of a given TET/JBP gene. All uncharacterized proteins in the neighborhood were then subjected to sequence and structure analyses (above) to determine their domain architectures and conserved features. Resampling of gene neighborhoods was performed separately for each genome using a jackknife-like procedure. The Perl script first removes all transposases and TET/JBP genes from the genome and then randomly reassigns each gene to an intergenic region of a chromosome/contig. This was repeated 10,000 times for each genome. The probability of finding at least as many TET/JBP-transposase pairs located within 10,000 bp of each other as observed in the real genomes was then estimated from the simulation.

Phylogenetic Trees and Congruence Comparisons. Maximum-likelihood phylogenetic trees were constructed using FastTree 2.1 (which implements an approximate method), MEGA5, and PHYML (42). Compare2trees was used to calculate the common tree topology between trees of TET/JBP and KDZ transposases (25). The TOPD/FMTS program was used for a nodal comparison between trees and for a randomization analysis to test whether the similarity between two trees is due to chance (26).

ACKNOWLEDGMENTS. This work was funded by the Intramural Research Program of the US Department of Health and Human Services, National Institutes of Health, National Library of Medicine (L.M.I., D.Z., and L.A.).

1. Aravind L, Anantharaman V, Zhang D, de Souza RF, Iyer LM (2012) Gene flow and biological conflict systems in the origin and evolution of eukaryotes. *Front Cell Infect Microbiol* 2:89.
2. Kehrer-Sawatzki H, Cooper DN (2007) Structural divergence between the human and chimpanzee genomes. *Hum Genet* 120(6):759–778.
3. Kazazian HH, Jr. (2004) Mobile elements: Drivers of genome evolution. *Science* 303(5664):1626–1632.
4. Iyer LM, Anantharaman V, Wolf MY, Aravind L (2008) Comparative genomics of transcription factors and chromatin proteins in parasitic protists and other eukaryotes. *Int J Parasitol* 38(1):1–31.
5. Aravin AA, Hannon GJ, Brennecke J (2007) The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. *Science* 318(5851):761–764.
6. Chalker DL, Yao MC (2011) DNA elimination in ciliates: Transposon domestication and genome surveillance. *Annu Rev Genet* 45:227–246.
7. Fedoroff NV (2012) Presidential address. Transposable elements, epigenetics, and genome evolution. *Science* 338(6108):758–767.
8. Shaheen M, Williamson E, Nickoloff J, Lee SH, Hromas R (2010) Metnase/SETMAR: A domesticated primate transposase that enhances DNA repair, replication, and decatenation. *Genetica* 138(5):559–566.
9. Kapitonov VV, Jurka J (2005) RAG1 core and V(D)J recombination signal sequences were derived from Transib transposons. *PLoS Biol* 3(6):e181.
10. Motl JA, Chalker DL (2009) Subtraction by addition: Domesticated transposases in programmed DNA elimination. *Genes Dev* 23(21):2455–2460.
11. Vogt A, Goldman AD, Mochizuki K, Landweber LF (2013) Transposon domestication versus mutualism in ciliate genome rearrangements. *PLoS Genet* 9(8):e1003659.
12. Iyer LM, Tahiliani M, Rao A, Aravind L (2009) Prediction of novel families of enzymes involved in oxidative and other complex modifications of bases in nucleic acids. *Cell Cycle* 8(11):1698–1710.
13. Tahiliani M, et al. (2009) Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* 324(5929):930–935.
14. Pastor WA, Aravind L, Rao A (2013) TETonic shift: Biological roles of TET proteins in DNA demethylation and transcription. *Nat Rev Mol Cell Biol* 14(6):341–356.
15. Iyer LM, Zhang D, Burroughs AM, Aravind L (2013) Computational identification of novel biochemical systems involved in oxidation, glycosylation and other complex modifications of bases in DNA. *Nucleic Acids Res* 41(16):7635–7655.
16. Borst P, Sabatini R (2008) Base J: Discovery, biosynthesis, and possible functions. *Annu Rev Microbiol* 62:235–251.
17. Mills RE, Bennett EA, Iskow RC, Devine SE (2007) Which transposable elements are active in the human genome? *Trends Genet* 23(4):183–191.
18. Beauregard PB, Guérin R, Turcotte C, Lindquist S, Rokeach LA (2009) A nucleolar protein allows viability in the absence of the essential ER-residing molecular chaperone calnexin. *J Cell Sci* 122(Pt 9):1342–1351.
19. Böhne A, et al. (2012) Zisupton—A novel superfamily of DNA transposable elements recently active in fish. *Mol Biol Evol* 29(2):631–645.
20. Dyda F, Chandler M, Hickman AB (2012) The emerging diversity of transpososome architectures. *Q Rev Biophys* 45(4):493–521.
21. Richardson JM, Colloms SD, Finnegan DJ, Walkinshaw MD (2009) Molecular architecture of the Mos1 paired-end complex: The structural basis of DNA transposition in a eukaryote. *Cell* 138(6):1096–1108.
22. Iyer LM, Aravind L (2012) ALOG domains: Provenance of plant homeotic and developmental regulators from the DNA-binding domain of a novel class of DIRS1-type retroposons. *Biol Direct* 7:39.
23. Li SJ, Hochstrasser M (1999) A new protease required for cell-cycle progression in yeast. *Nature* 398(6724):246–251.
24. Koonin EV, Dolja VV (1993) Evolution and taxonomy of positive-strand RNA viruses: Implications of comparative analysis of amino acid sequences. *Crit Rev Biochem Mol Biol* 28(5):375–430.
25. Nye TM, Liò P, Gilks WR (2006) A novel algorithm and web-based tool for comparing two alternative phylogenetic trees. *Bioinformatics* 22(1):117–119.
26. Puigbò P, Garcia-Vallvé S, McInerney JO (2007) TOPD/FMTS: A new software to compare phylogenetic trees. *Bioinformatics* 23(12):1556–1558.
27. Rountree MR, Selker EU (2010) DNA methylation and the formation of heterochromatin in Neurospora crassa. *Heredity (Edinb)* 105(1):38–44.
28. Freedman T, Pukkila PJ (1993) De novo methylation of repeated sequences in Coprinus cinereus. *Genetics* 135(2):357–366.

29. Zemach A, McDaniel IE, Silva P, Zilberman D (2010) Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science* 328(5980):916–919.
30. Iyer LM, Abhiman S, Aravind L (2011) Natural history of eukaryotic DNA methylation systems. *Prog Mol Biol Transl Sci* 101:25–104.
31. Stajich JE, et al. (2010) Insights into evolution of multicellular fungi from the assembled chromosomes of the mushroom Coprinopsis cinerea (Coprinus cinereus). *Proc Natl Acad Sci USA* 107(26):11889–11894.
32. Duplessis S, et al. (2011) Obligate biotrophy features unraveled by the genomic analysis of rust fungi. *Proc Natl Acad Sci USA* 108(22):9166–9171.
33. Zolan ME, Heyler NK, Stassen NY (1994) Inheritance of chromosome-length polymorphisms in Coprinus cinereus. *Genetics* 137(1):87–94.
34. Pukkila PJ, Skrzynia C (1993) Frequent changes in the number of reiterated ribosomal RNA genes throughout the life cycle of the basidiomycete Coprinus cinereus. *Genetics* 133(2):203–211.
35. Wray J, et al. (2010) The transposase domain protein Metnase/SETMAR suppresses chromosomal translocations. *Cancer Genet Cytogenet* 200(2):184–190.
36. Altschul SF, et al. (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res* 25(17):3389–3402.
37. Eddy SR (2009) A new generation of homology search tools based on probabilistic inference. *Genome Inform* 23(1):205–211.
38. Edgar RC (2004) MUSCLE: A multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5:113.
39. Söding J, Biegert A, Lupas AN (2005) The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res* 33(Web Server issue): W244–W248.
40. Cuff JA, Clamp ME, Siddiqui AS, Finlay M, Barton GJ (1998) JPred: A consensus secondary structure prediction server. *Bioinformatics* 14(10):892–893.
41. Finn RD, et al. (2010) The Pfam protein families database. *Nucleic Acids Res* 38(Database issue):D211–D222.
42. Price MN, Dehal PS, Arkin AP (2010) FastTree 2—Approximately maximum-likelihood trees for large alignments. *PLoS ONE* 5(3):e9490.